DOCUMENT RESUME

ED 429 101 TM 029 639

AUTHOR Inoue, Yukiko
TITLE Evaluating Intelligent Tutoring Systems.
PUB DATE 1999-00-00
NOTE 11p.
PUB TYPE Information Analyses (070)
EDRS PRICE EDRS Price MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS *Computer Managed Instruction; Educational Research; Evaluation Methods; *Intelligent Tutoring Systems; Interaction; Literature Reviews; Meta Analysis; Research Methodology; *Tutorial Programs

ABSTRACT

Although there are three main computer-assisted instruction (CAI) formats (drill and practice, simulation, and tutorial), tutorial format is particularly suitable to conceptual school subjects because this format: (1) explains the concept of the text; (2) tests the understanding of the topic; and (3) selects the next step of the subject. When a knowledge module is added to the tutorial format, it is called an intelligent tutoring system (ITS). In short, an intelligent computer-assisted instruction (I-CAI) does emphasize interactions of the learner with the computerized system. In light of methodology and interpretation, this paper examines ITS evaluation studies conducted by leading researchers in the ITS field. Although ITS evaluations greatly influence interest in and support for future ITS work, valid evaluation methodologies are lacking mainly because the system itself is fairly new. The critical component in ITS is its knowledge component, which may not have been evaluated adequately to verify the knowledge. ITS researchers, therefore, have simply suggested using procedures used in evaluation studies of artificial intelligence or expert systems. Such procedures, however, may or may not be applicable in educational settings, and therefore, further research is strongly recommended. Summative evaluations were found to be more difficult to execute than formative evaluations because summative evaluations would involve the comparison of ITS with human tutors using traditional teaching methods across extensive problem domains. (Contains 18 references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 429 101

EVALUATING INTELLIGENT TUTORING SYSTEMS

Yukiko Inoue, Ph.D.

The University of Guam
Educational Research, College of Education
UOG Station, Mangilao, Guam 96923
Phone: 671- 735-2447; Fax: 671-734-3651
Email: yinoue@uog.edu

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY HAS BEEN GRANTED BY
Yukiko Inoue
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2A

TM029639

BEST COPY AVAILABLE

EVALUATING INTELLIGENT TUTORING SYSTEMS

Yukiko Inoue, Ph.D.
College of Education, University of Guam

ABSTRACT

Although there are three main computer-assisted instruction (CAI) formats (drill and practice, simulation, and tutorial), tutorial format is particularly suitable to conceptual school subjects because this format (1) explains the concept of the text, (2) tests the understanding of the topic, and (3) selects the next step of the subject. When knowledge module is added to the tutorial format, it is called an intelligent tutoring system (ITS). In short, an intelligent computer-assisted instruction (I-CAI) does emphasize interactions of the learner with the computerized system. In the light of methodology and implementation, this paper examined ITS evaluation studies conducted by leading researchers in the ITS field. Albeit ITS evaluations greatly influence both interest in and support for the future ITS work, valid evaluation methodologies are lacking mainly because ITS itself is fairly new. The critical component in ITS is its knowledge component that has not been evaluated adequately to verify the knowledge. ITS researchers, therefore, simply suggested employing procedures used in evaluation studies of artificial intelligent or expert system. Such procedures, nevertheless, may or may not be applicable in educational settings and therefore further research is greatly recommended. Finally, summative evaluations were found to be more difficult than formative evaluations to execute since they would involve the comparison of ITS with human tutors rousing traditional teaching methods across the extensive problem domains.

Computer-assisted instruction (CAI) is not a new technique; in fact, computers have been used in educational settings since the late 1950s. The field of CAI covers all kinds of instructional systems in which computers are used to support teaching and learning. If properly designed and implemented, CAI has many advantages: For instance, CAI offers an interactive response, immediate feedback, infinite patience, and motivation, as well as an ability to maintain accurate records of student progresses (Sloane, Gordon, Gunn, & Mickelsen, 1989). Thus CAI will be a powerful complement to traditional instruction. Although there are three main CAI formats (drill and practice, simulation, and tutorial), tutorial is particularly suitable to conceptual school subjects since it (1) explains the concept of the text, (2) tests the understanding of the topic, and (3) selects the next step of the subject (Kearsley, 1993). When knowledge module is added to tutorial format, it is called an intelligent-tutoring-system (ITS). In short, an intelligent computer-assisted instruction (I-CAI) emphasizes interactions of the learner with the computerized system (Alessi & Trollip, 1991).

Albeit ITS evaluation studies greatly influence interest in and support for the future ITS work (Winnie, 1993), valid evaluation methodologies are lacking since ITS itself is fairly

new (Mark & Greer, 1993; Murray, 1993; Shute & Regian, 1993). In the light of methodology and implementation, therefore, this paper examined the state of the art of ITS evaluations. The paper consists of the following four sections: (1) the aspects of the ITS evaluation, (2) the methodological issues of the ITS evaluation, (3) the selected examples of the ITS evaluation, and (4) summary and conclusion.

Aspects of the Intelligent Tutoring System Evaluation

Evaluation in education refers to a systematic effort to gather and to interpret information with which to assess educational values and teaching methods. ITS is such an emerging technology that has no accepted cadre of formal research methodologies and therefore it needs to be evaluated. In brief, evaluations sway public and expert opinions about ITS and what people portray as progress in education.

This paper thus attempted to shade some light on ITS work; furthermore, the methodology and theory of the ITS evaluation are virtually coupled. In the area of methodology, in particular, the following four aspects are concerned: (1) summative versus formative, (2) internal versus external, (3) quantitative versus qualitative, and (4) formal versus informal. Summative evaluations are carried out to support formal claims about outcomes associated with completed systems. Formative evaluations, on the other hand, refer to the evaluation of instructional system while the system is being developed. Internal evaluations relate the impact of the ITS's architecture to the system behavior, whereas external evaluations address the educational impact of ITS on the learner. Quantitative evaluations refer to the approach of using rigorous, objective, and numerical analyses, whereas qualitative evaluations use subjective approaches involving attitudes and perceptions. Lastly, quantitative approaches are considered formal, whereas qualitative approaches are usually informal.

Methodological Issues of the Intelligent Tutoring System Evaluation

Methodologies of the ITS Evaluation

In the matrix of the framework for ITS, which is based on a distinction between formative and summative evaluations, Mark and Greer (1993) have evaluated the suitability of valuational techniques against ITS components and have suggested to use methods from artificial intelligence or CAI because of few methods developed for the ITS evaluation. The issue in evaluating ITS is the evaluation of the knowledge embedded in ITS; therefore, the

knowledge of evaluation methods from artificial intelligence can be used. The framework is united to learning strategies that are important in the conscious monitoring of human information processing (Baron & Sternberg, 1987). The above matrix judges the evaluation of ITS's architecture (internal) but the educational impact (external) is more difficult to assess.

Evaluations and Designs for ITS

Both concept and methodology in evaluation and design for ITS are complementary and quite similar in essence. Although any study on the combination of the evaluation and the design has not been found, Rowland's (1994) generic exposition on the design and evaluation of educational system can be used as a guideline for the relationship between these areas. One vital objective of conceptual studies is the integration of the design and the evaluation (these two areas are often portrayed as separate entities). It can be possible that such integrations combine problem-focus with solution-focus and rationality with creativity.

Principles of Summative Evaluations for the ITS Efficiency

One of the essential goals of evaluation systems is to measure the ITS efficiency. Summative evaluations involve experimental designs in which an appropriate balance between external validity and internal validity is determined by the three steps: (1) compare evaluation studies, (2) codify the process of designing and conducting competent evaluation studies, and (3) maximize the utility of future ITS evaluation studies. Shute and Regian (1993) have maintained the following principles of summative evaluations for the ITS efficiency:

1) Delineate the goals of ITS (prior to the design of evaluation).
2) Define the goals of the evaluation study.
3) Select the appropriate design to meet the defined goals.
4) Instantiate the design with the appropriate measure, number and type of subjects, and control conditions.
5) Make logistical preparations for conducting the evaluation study.
6) Conduct a pilot-test for the ITS (tutor) and the study.
7) Determine the primary data analyses as plan the study.

While these principles may sound as generic principles for any experimental design, their details embrace adaptations for ITS. Building tutoring systems without evaluating them is a complete waste. Although it is not touched upon in the principles, the use of appropriate measurement instruments might be crucial.

Quantitative Evacuations for ITS

As noted by Twidale (1993), there are many potential problems in conducting quantitative evaluations for ITS. Such problems, however, can be minimized with apropos procedures, just as Legree, Gillis, and Orey (1993) have tried to associate different evaluations with potential problems using their own procedures that examined the diagnostic accuracy and the knowledge-based reliability. Estimating the external effectiveness of ITS requires comparing ITS with human beings and teaching methods in different situations, but evaluating the internal component of ITS requires the procedure that assesses the instructional validity of the component. The procedure was applied by Legree et al. (1993) to the three different types of ITS. The results indicated that traditional evaluation methods were useful for estimating the overall effectiveness of ITS; however, ITS could be improved with special evaluation methods of internal components.

Formative Qualitative Evaluations for the Exploratory ITS Study

The aim of exploratory research is to identify key issues and tradeoffs rather than to challenge or to extend existing methods which, according to Kearsley (1993), is the most difficult step in building tutoring systems. Murray (1993) has represented the five evolution and design paradigms for the exploratory ITS studies: (1) formative evaluation, (2) qualitative evaluation, (3) iterative design, (4) participatory design, and (5) case study. In the exploratory area of knowledge acquisition, it is viable to control only few variables but not viable to use the sufficient sample size and the appropriate statistical method. That is, qualitative evaluations are appropriate for the exploratory area, even though qualitative evaluations are subjective and descriptive rather than conclusive.

Informal Evaluations for ITS

Students' interactions with ITS are likely to involve unexpected and unpredictable aspects that can only be revealed when frequent uses of the informal in-depth study on the prototype of ITS in the development process. Twidale (1993) made this observation as an example of advantages of informal evaluation for revealing problems with ITS while it was being developed. This approach, indeed, complements and substitutes rigorous controlled evaluations. While formal methods are suitable for summative evaluations, informal methods are suitable for formative evaluations. If informal evaluations are not applicable, well-

controlled summative experiments are the best means of the ITS evaluation.

Cognitive Mediation Perspectives of the ITS Evaluation

A unique evaluation method has been proposed by Or-Bach and Bar-On (1993), and they have argued that ITS should be evaluated from the cognitive mediative perspective rather than from the computer system perspective since internal evolutions are more productive avenues to follow. This method focuses not on the measurable outcomes of instructional process but on what is provided by systems to construction (learning). Unlike Mark and Greer (1993) who have inspected the components of ITS, Or-Bach and Bar-On (1993) have maintained that evaluations should consider different dimensions of the mediation process. The above model, which is called TALK (Tactics, Analogy, Language, Knowledge), is to be applied to mathematics for optimal solutions by the step-by-step process. While the proposal is unique, its implementation is not easy because of its high cost.

Integrating Cognitive Theories Into ITS

Principles of behavioral theories have been used to improve CAI. Recent developments in both cognitive and I-CAI theories have suggested a possible merger of these areas (e.g., see Hannafin & Rieber, 1991). Oray and Nelson (1991) have developed a general model of ITS and have suggested the ways to improve ITS using design principles derived from researchers in cognitive science. The authors have identified the four components in ITS: (1) the user interface, which provides a means of the two-way communication in which the learner is engaged in activity while the system interprets the activity to make meaningful responses; (2) the learner model, which is a representation of misconception that occurs when the learner is exposed to the content; (3) the expert model, which is a database that represents the knowledge of an expert in the domain; and (4) the pedagogic knowledge, which receives the result of a comparison of the learner's knowledge with the knowledge in expert model.

In order for developers of ITS to integrate these findings into their practices, a new perspective is necessary, that is, different assumptions about learning, not assumptions in instructional design models currently in use (Oray & Nelson, 1991). The addition of cognitive approaches to ITS is a valuable contribution to ITS evaluations since the results derived from such evaluations can be employed to improve ITS.

Selected Examples of the ITS Evaluation

The Effectiveness of Intelligent Multimedia Systems

Computer-based multimedia (such as hypertext and CD-ROM) are popular for many reasons (for example, the increased availability of software, the reduced cost of hardware, and the availability of Internet interfaces). Researches on computer-based multimedia in education are criticized (Rosenberg, 1987) as being designed in careless manners in respect of the effect of instruction and learning context. Livergood (1993-1994) compared the effectiveness of hardcopy materials, regular computer-based multimedia, and intelligent multimedia and has found that adding multimedia contents to traditional instructional materials did not invariably improve students' performance in computer science (Livergood's subjects were college students): However, significant improvement was gained when intelligence multimedia was utilized. Some strong points of the study above are the control of potentialities influencing variables and the large sample sizes (41, 73, and 95 respectively). The drawbacks include that the results may be limited to one set of the domain knowledge.

Effect Sizes of I-CAI

Effect size (the minimum treatment effect) is the mean difference between the experimental and control groups divided by the standard deviation of the control group. Although the synthesis of effect size is the subject of meta-analysis, the larger the size is, the more superior the research will be. The meta-analysis of CAI studies by Kulik and Kulik (1991) has shown the effect size of .30. Despite the high hope for CAI, prior studies have not established that CAI is more effective than other educational provisions. A study on ITS (based on AI) by Fraser and Teh (1994) involved 671 students in 12 high schools in Singapore in the achievement of geography (non-I-CAI versus I-CAI) in six weeks. The ITS resulted in significantly higher achievement with the large effect size of 3.5. In other words, Fraser and Teh's (1994) study is the well-designed evaluation of ITS and introduces a new dimension (psychological environment) into the future studies of CAI as well as ITS, even though it needs to be replicated due to the uniqueness of the Singapore milieu.

Evaluations of Commercial Software in Building ITS

One major problem with ITS is the extensive software development cost. The development time of ITS can be reduced with the use of standard and off-the-shelf hypermedia

packages. If the system is built with the standard package, is it as good as system custom-made with a C programming language? Orey and Nelson (1991) used the commercial hypermedia tool to develop the hypermedia-based ITS and compared it to the system developed by the C language. The domain was a military training course for the operation and the troubleshooting of communication equipments. As a result, the system developed with the commercial package required only about forty percent of time compared to the one required by custom programs. The results tell the possibility of breakthrough in the development of significantly more effective ITS, though they are limited to a specific software package and programming language.

Summary and Conclusion

It was attempted in this paper to cover most of the important studies conducted by leading researchers in the field of ITS. Dealing with an emerging educational field has such an advantage that the number of research publications is limited and relatively easy to cover most of the studies, though many of them are theoretical and exploratory in nature. To sum up, the ITS evaluation has reached a consensus regarding the following four points: (1) the ITS evaluation is vital but difficult because of the diversity of ITS; (2) the evaluation of the overall effectiveness of ITS can be conducted with traditional evaluation methods; (3) it is possible not only to evaluate the component of ITS while it is being developed but also to modify and to improve the system while it is being prototyped; and (4) the behavioral theory-based evaluation should be supplemented with the cognitive theory-based principles.

The critical component in ITS is its knowledge component, which has not been evaluated adequately to verify the knowledge, and therefore researchers have suggested employing the procedures used in the research of artificial intelligence and expert systems. Such procedures, nevertheless, may or may not be applicable in educational settings and further studies are greatly recommended in this area. Most of the discussion in this paper has centered around formative evolutions. In conclusion, summative evaluations are indeed more difficult to execute mainly because they involve the comparisons of ITS with human tutors rousing traditional teaching methods across the extensive problem domains.

REFERENCES

Alessi, S.M., & Trollip, S. R. (1991). Lesson evaluation and revision. Computer-based instruction: Methods and development (2nd Ed.). Englewood Cliffs, NJ: Prentice-Hall.

Baron, J.B., & Sternberg, R. J. (1987). Teaching thinking skills: Theory and proactive. New York: H. Freeman.

Fraser, B.J., & Teh, G. P.L. (1994). Effect size associated with micro-prolog-based computer-assisted learning. Computers and Education, 23(3), 187-196.

Hannafin, M. J., & Rieber, L.P. (1989). Psychological foundation of instructional technologies: Part 1. Educational Technology Research and Development, 37, 91-101.

Kearsley, G. (1993). Computer-based training. Reading, MA: Addison-Wesley.

Kulick, C.C., & Kulick, J.A. (1991). Effectiveness of computer-based instruction: An updated analysis. Computer and Human Behavior, 71, 75-94.

Legree, P.J., Gillis, P.D., & Orey, M.A. (1993). The quantitative evaluation of intelligent tutoring system applications: Product and process criteria, Journal of AI Education, 4(2/3), 209-226.

Livergood, N.D.(1993-1994). A study of the effectiveness of a multimedia intelligent tutoring system. Journal Educational Technology Systems, 22(4), 337-344.

Mark, M.A., & Greer, J.E.(1993). Evaluation methodologies for intelligent tutoring systems. Journal of AI in Education, 4(2 & 3), 129-153.

Murray, T. (1993). Formative qualitative evaluation for exploratory ITS research. Journal of AI in Education, 4(2 & 3), 179-207.

Or-Bach, R., & Bar-On, E. (1993). "TALK"in about evaluation, Journal of AI in Education, 4(2 & 3), 227-243.

Orey, M.A., & Nelson, W.A. (1991, February). Using intelligent tutoring design principles to integrate cognitive theory into computer-based instruction. Paper presented at the annual convention of the association for educational communication and technology, Orlando, FL.

Rosenberg, R. (1987, November). A critical analysis of research on intelligent tutoring systems. Educational Technology, pp. 7-13.

Rowland, G. (1994). Designing and evaluating: Creating futures and appreciating error, Educational Technology, 34(1), 10-22.

Shute, V. J., & Regian, J. W. (1993). Principles of revaluating intelligent tutoring systems. Journal of AI in Education, 4(2 & 3), 245-271.

Sloane, H.N., Gordon, H.M., Gunn, C., & Mickelsen, V.G. (1989). Evaluating Educational Software: A Guide for Teaches. Englewood Cliffs, NJ: Prentice Hall.

Twidale, M. (1993). Redressing the balance: The advantages of informal evaluation techniques for intelligent learning environments. Journal of AI in Education, 4(2 & 3), 155-178.

Winnie, P.H. (1993). A landscape of issues in evaluating adaptive learning systems. Journal of AI in Education, 4(4), 309-332.